

BIONUMERICS Tutorial:

Importing and processing VNTR sequencer curve files

1 Aim

Comprehensive tools for processing electrophoresis fingerprints, both from slab gels and capillary sequencers are incorporated into BIONUMERICS. When fingerprints are run on a capillary sequencers the resulting data can have two different formats: curve files (also referred to as electropherograms, chromatogram files or trace files), or peak tables. In this tutorial we will focus on the import of curve files. Since curve files contain the raw data, the fingerprint processing steps (e.g. normalization, band assignment, ...) are also covered in this tutorial.

2 Sample data

Raw curve files from Applied Biosystems and Beckman (now AB SCIEX) sequencers can be imported directly in BIONUMERICS. A batch of Applied Biosystems curve files can be downloaded from the Applied Maths website (https://www.applied-maths.com/download/sample-data, click on "VNTR sequencer trace files"). These example files will be used to illustrate the import steps. Eight VNTR targets per sample were amplified: two pools were generated (MP1 and MP2), each containing four PCR products, using four color dyes (6-FAM, VIC, NED, and PET).

3 Importing curve files

- 1. Create a new database (see tutorial "Creating a new database") or open an existing database.
- 2. Select *File* > *Import...* (, Ctrl+I) to call the *Import* dialog box, choose *Import curves* under *Fingerprint type data* and press <*Import*>.
- 3. Browse to the downloaded and unzipped example data folder VNTR sequencer trace files and select all Applied Biosystems curves files (extension .fsa). Press < **Open**>.

The files are displayed in the *Input* wizard page and the default suggested *Fingerprint file name* is the folder name.

4. Change the name to **Batch1** and press <*Next*> (see Figure 1).

Input			
Select the data t	o import.		
Select curve file(s):	C:\Users\10\N2006-0003 MP1-140708.fsa A Browse		
	C:\Users\10\N2006-0003_MP2-140708.fsa	•	
	C:\Users\10\N2006-0004_MP1-140708.fsa	All	
	C:\Users\10\N2006-0021_MP1-140708.fsa		
	C:\Users\10\N2006-0021_MP2-140708.fsa		
	C:\Users\10\N2006-0022_MP1-140706.fsa		
	C:\Users\10\N2006-0040_MP1-140708.fsa		
	"C:\Users\10\N2006-0040_MP2-140708.fsa ♥ 40 file(s), 3 Mb		
Fingerprint file name:	Batch1		

Figure 1: Selected curve files and fingerprint file name.

The way the information should be imported in the database should be specified with an import template. In the example data set, the dye name can be parsed from the file information and the sample and pool information can be parsed from the file names.

5. Select the predefined template *Example import with pools* and press the *<Preview>* button to check the rules applied on the selected files.

revie	eview						
Nr.	Fingerprint pool	Fingerprint dye	Кеу		^		
1	MP1-140708	6-FAM	N2006-0003				
2	MP1-140708	VIC	N2006-0003				
3	MP1-140708	NED	N2006-0003				
4	MP1-140708	PET	N2006-0003				
5	MP1-140708	LIZ	N2006-0003				
6	MP2-140708	6-FAM	N2006-0003				
7	MP2-140708	VIC	N2006-0003				
8	MP2-140708	NED	N2006-0003				
9	MP2-140708	PET	N2006-0003				
10	MP2-140708	LIZ	N2006-0003				
11	MP1-140708	6-FAM	N2006-0004				
12	MP1-140708	VIC	N2006-0004				
13	MP1-140708	NED	N2006-0004				
14	MP1-140708	PET	N2006-0004				
15	MP1-140708	LIZ	N2006-0004				
16	MP2-140708	6-FAM	N2006-0004				
17	MP2-140708	VIC	N2006-0004				
18	MP2-140708	NED	N2006-0004				
19	MP2-140708	PET	N2006-0004		~		

Figure 2: Preview of predefined template.

In the preview we see that the part of the file name appearing before the underscore is linked to the *Key* field. The curve dyes are also correctly linked to the *Fingerprint dye* type. In a next step, we will modify the rule for the *Fingerprint pool*: we will link the part of the file name appearing after the underscore and before the hyphen ("-") to the *Pool* field.

- 6. Close the preview.
- 7. Make sure the predefined template *Example import with pools* is still selected and press the <*Edit*> button to edit the rules.

- 8. Check the option *Show advanced options*, make sure the second row is selected in the grid panel and press the *<Edit parsing>* button.
- 9. In the *Data parsing* dialog box, update the data parsing string: "*_[DATA]-*". The asterisk will serve as wildcard.
- 10. Press the *Preview* button and press *OK* when the parsing is correct (see Figure 3).

Edit data pa	arsing			?	×
Parse	component	: find the componen	t '[DATA]', use ''	" as wildcar	d
🔿 Regula	ır expressi	on: match the expre	ssion and use t	he subexpre	ssion
Data parsin	g string:	*_[DATA]-*		~	
Data decora	ation:	[DATA]		~	
Preview					
Data:	N2006-00	003_MP1-140708	Preview	N	
Output:	MP1				
			ОК	Can	cel

Figure 3: Parsing string.

The Import rules dialog box is updated.

11. Press the *Preview* button to check the outcome of all defined import rules (see Figure 4).

Nr.	Fingerprint pool	Fingerprint dye	Key	
1	MP1	6-FAM	N2006-0003	
2	MP1	VIC	N2006-0003	
3	MP1	NED	N2006-0003	
4	MP1	PET	N2006-0003	
5	MP1	LIZ	N2006-0003	
6	MP2	6-FAM	N2006-0003	
7	MP2	VIC	N2006-0003	
8	MP2	NED	N2006-0003	
9	MP2	PET	N2006-0003	
10	MP2	LIZ	N2006-0003	
11	MP1	6-FAM	N2006-0004	
12	MP1	VIC	N2006-0004	
13	MP1	NED	N2006-0004	
14	MP1	PET	N2006-0004	
15	MP1	LIZ	N2006-0004	
16	MP2	6-FAM	N2006-0004	
17	MP2	VIC	N2006-0004	
18	MP2	NED	N2006-0004	
19	MP2	PET	N2006-0004	_

Figure 4: Preview of the import rules.

12. Close the preview and press < *Next* > to go to the next step.

In the example data set, the LIZ channel contains the size standard (GeneScan 500 LIZ).

13. Make sure *LIZ* is selected as *Reference dye* and press <*Next*> and <*Finish*>.

The predefine import template is now updated.

14. Make sure <*Create new*> is selected as fingerprint type experiment, make sure the *Example import with pools* template is selected and press <*Next*>.

- 15. Specify a name for the new base fingerprint type experiment (e.g. **MLVA**) and press < *OK* > and confirm.
- 16. Confirm the proposed OD range.

Create new	?	×
Provide a name for the new f	fingerprint	type:
ОК	Car	ncel

Figure 5: New base fingerprint type experiment.

A fingerprint type needs to be present in the database for each pool and dye combination. The names of these fingerprint types are composed of the base fingerprint type name, followed by the pool name, and the name of the dye. If one or more of these fingerprint types are not present in the database, a new dialog box pops up, listing the missing fingerprint types (see Figure 6).

MLVAMP16-FAM		
MI VAMP11 IZ		
MLVAMP1NED		
MLVAMP1PET		
MLVAMP1VIC		
MLVAMP26-FAM	\mathbf{v}	

Figure 6: Missing fingerprint experiments.

- 17. Confirm the creation of the missing fingerprint type experiments.
- 18. Press <*Next*> to confirm the creation of new entries in the database.
- 19. Make sure *Open curve preprocessing window* is checked in the last step and press <*Finish*>.

For each dye checked in the *Dyes panel* of the *Import data* dialog box, a new fingerprint file is created, composed of the file name specified and the name of the dye (e.g. Batch1_LIZ). These files are displayed in the *Fingerprint files* panel. The reference file is shown in the **Link** column. Double-clicking on a fingerprint file opens the *Fingerprint* window. If lane information was imported with the individual lanes, this information is displayed in the *Fingerprint information panel*.

The imported fingerprint lanes are linked to new entries in the database. The lanes are linked to the corresponding fingerprint "dye" type. The names of these fingerprint types are composed of the base fingerprint type name, followed by the pool name, and the name of the dye. The fingerprint type experiments are displayed in the *Experiment types* panel.

Entries for which fingerprint data was imported are selected in the database.

After data import, the *Main* window looks as in Figure 7.

eriment t	types			Database entries			Comparisons
易	+ 🖒 🛛 🗞	£ v.	↑ ↓ <all experiment<="" th=""><th>\$1 + P (</th><th>8 €, Ē</th><th><all entries=""> し</all></th><th>+ 13 ⊗ €. 61 ⊽</th></all>	\$1 + P (8 €, Ē	<all entries=""> し</all>	+ 13 ⊗ €. 61 ⊽
#	Name	Туре	-	Key	Modified date	🖵 1 2 3 4 5 6 7 8 9 10 11	Name Modified da
	1 MLVA	Fingerprint types	^	N2006-0003	2020-04-20 11:00:02		
	2 MLVAMP16-FAM	Fingerprint types		N2006-0004	2020-04-20 11:00:02		
	3 MLVAMP1LIZ	Fingerprint types		N2006-0021	2020-04-20 11:00:02		
	4 MLVAMP1NED	Fingerprint types		N2006-0022	2020-04-20 11:00:02		
	5 MLVAMP1PET	Fingerprint types		N2006-0040	2020-04-20 11:00:02		
	6 MLVAMP1VIC	Fingerprint types		N2006-0063	2020-04-20 11:00:02		
	7 MLVAMP26-FAM	Fingerprint types		N2006-0082	2020-04-20 11:00:02		<
	8 MLVAMP2LIZ	Fingerprint types		N2006-0099	2020-04-20 11:00:02		`
	9 MLVAMP2NED	Fingerprint types		N2006-0112	2020-04-20 11:00:02		Identification projects Decision networks
1	10 MLVAMP2PET	Fingerprint types		N2006-0123	2020-04-20 11:00:02		
1	11 MLVAMP2VIC	Fingerprint types		N2006-0152	2020-04-20 11:00:02		2 + 🗗 ⊗ 🔩 🛍
			~	N2006-0170	2020-04-20 11:00:02		Name Modified da
				N2006-0172	2020-04-20 11:00:02		
ields E	Database design			N2006-0179	2020-04-20 11:00:02		
F3		⊽ ↑ ↓	<all entry="" fields=""></all>	N2006-0185	2020-04-20 11:00:02		
				N2006-0186	2020-04-20 11:00:02		
Nam	ne <mark>Field</mark>	type	•	N2006-0208	2020-04-20 11:00:02		
			^	N2006-0231	2020-04-20 11:00:02	• • • • • • • • • • •	
			*	N2006-0232	2020-04-20 11:00:02	• • • • • • • • • • • •	<
print files	S Power assemblies An	notations		N2006-0233	2020-04-20 11:00:02		
							Alignments BLAST projects Chrom. Comp.
+		d ∇ <all f<="" td=""><td>Fingerprint files></td><td></td><td></td><td></td><td></td></all>	Fingerprint files>				
File na	ame Experiment	type Link	Modified date				
Batch1	1_LIZ MLVA		2020-04-20 11:07:40				Name Modified da
	1_6-FAM MLVA	Batch1_LIZ	2020-04-20 11:07:40				
Batch1		Batch1 LIZ	2020-04-20 11:07:40				
Batch		Batch1_LIZ	2020-04-20 11:07:40				
Batch		Batch1_LIZ	2020-04-20 11:07:40				
		_uuuii_uu					
			V				

Figure 7: The Main window after import of the data.

4 Processing curve files

When the option *Open curve preprocessing window* was checked in the last step of the import routine, the *Fingerprint curve processing* window opens when pressing the *<Finish*> button. All channels from the run are automatically loaded and displayed in the *Fingerprint curve processing* window.

The *Fingerprint curve processing* window can also be called from the *Main* window by highlighting one of the channels in the *Fingerprint files* panel and selecting *Open fingerprint data...* (\blacksquare). Alternatively, you can first open the *Fingerprint* window with *Edit* > *Open highlighted object...* (\boxdot , Enter) and then select *File* > *Edit fingerprint data...* (\blacksquare).

1. Click on the onicon left of the data channels in the Channels panel.

The data channels are now hidden from the view and its icons are displayed as .

2. Use the zoom sliders on the left and on top to optimize the display of the fingerprint curves.

Since the raw chromatogram files have not undergone any preprocessing, normalization will have to be performed. This requires a *reference system* to be defined, based upon the marker peaks available in the reference dye.

- 3. Make sure the reference dye is the only dye visible in the upper panel.
- 4. Select *Bands* > *Search reference bands...* (41, Ctrl+F) to call the *Search reference bands* dialog box.
- 5. Leave the default settings unaltered and press < OK >.

The bands that fall within the specified criteria are marked with a solid line at the band's position (see Figure 8).

6. To have a reference system automatically created based on a lane containing commercial size marker, first highlight a suitable lane and then select *References* > *Define size standard...*.

Finger	prints											
Q												
24 t No	ands on highlighted curve (75 aand selected	62 pts)	No reference points define	ł				No re	ference points	s defined		
r.			, the			1 I	I	11	Ī	I II		^
■ N:	2006-0123 [MLVAMP1LIZ]		Ť Ť Ť Ľ +		TTT	J T	Ţ					
	2006-0123 [MLVAMP2LIZ]							<u> </u>				
I N	2006-0152 [MLVAMP1LIZ]											~
Chann	els		Overview									
8					55	-335-	5	5	5 5	5 5 -	5 45	
	File name	Channel	-	3	33	- 225 -	3	3	2, 2	2 - 3 -	$\beta_{i} = \beta_{i}$	
	Batch1_LIZ	LIZ	^		(3, 3)	- 555,	- 5,	- i,	- 1, - 1 - 1	$i_{j_1} = i_{j_1}$	3 <u>3</u> -	
	Batch1_6-FAM	6-FAM			TT.	-111	1	1	1 L			
	Batch1_VIC	VIC		1900	11	- 555	5	5	5 5	5 5	5 5	
\odot	Batch1_NED	NED		1	- { - {	- <u>1,6</u>	1	1	5, 5	5. S.	1, 1 <u>1</u> ,	
© <	Batch1_PET	PET	>	3	11	335	1	1	1 11	~ <i>f</i> '	1 11	

Figure 8: The *Fingerprint curve processing* window only displaying the reference dye.

This will display the *Size standard* dialog box, from which a size marker can be selected (see Figure 9).

Size standard	?	×	<
Size standard Pre-defined standard: GeneScan™ 120 LIZ® GeneScan™ 500 LIZ® GeneScan™ 1200 LIZ® GeneScan™ 400HD ROX™ GeneScan™ 400HD ROX™ GeneScan™ 500 ROX™ GeneScan™ 500 ROX™ GeneScan™ 500 XL ROX™ GeneScan™ 500 XL ROX™ GeneScan™ 500 XL RAMRA™ GeneScan™ 500 XL TAMRA™ GeneScan™ 720 LIZ®	? Standard fragment list: 35,50,75,100,139,150,160,200,250,300,340,350,400,45 0,500 Invert ordering Map to selected lane Right to left Left to right Pattern match		
	ОК	Cancel	

Figure 9: Select marker from the list.

In the example curve files, the GeneScan 500 LIZ size standard is used as reference, containing 16 bands with known molecular weight.

- 7. Select GeneScan 500 LIZ from the list, select Pattern match and press < OK > twice.
- 8. Save the data to the database with *File* > *Save* (E, Ctrl+S). Confirm the action.

BIONUMERICS will automatically create the reference system and calibration curve for each of the fingerprint types. Since this allows the calculation of metrics information, a metrics scale now becomes available in the upper part of the *Fingerprints* panel.

Normalization is achieved by assigning bands in the reference channel to external reference positions.

- 9. To normalize a complete run at once, select *Normalization* > *Auto assign reference positions (all lanes)...* (^(A), Ctrl+A), leave all settings unaltered and press <*OK*>.
- 10. When the assignment of the marker bands to reference positions is made, the data can be shown in normalized mode with *Normalization* > *Show normalized view* (善, Shift+N) (see Figure 10).

Fingerprints		
•		
23 bands on highlighted curve (7982 pts) No band selected		
I N2006-0003 [MLVAMP1LIZ]		^
C 10206-0003 [MLVAMP2LIZ]		
■ N2006-0004 [MLVAMP1LIZ]		
N2006-0004 [MLVAMP2LIZ]		¥
Channels	Overview	
8		
File name Channel	▼	
Batch 1_LIZ LIZ		
Batch 1_6-FAM 6-FAM		
Batch 1_VIC VIC		
Batch 1_NED NED		
Batch 1_PET PET		
	× 01	

Figure 10: Normalized view - reference dye.

- 11. Click on the icon left of the data channels in the Channels panel.
- 12. Click on the o icon left of the LIZ reference channel.

The data channels are now shown and the reference channel is hidden from the view.

- 13. Select *Bands* > *Search data bands...* (, Ctrl+Shift+F) to call the *Search data bands* dialog box.
- 14. Check *Remove doublets*, *Remove shadow bands*, *Filter by fragment length* and specify a minimum length of 50. Press <*OK*>.
- 15. Save the changes and close the *Fingerprint curve processing* window. Confirm the new defaults.

The reference system and calibration curve are automatically created for each of the fingerprint types. We can check this from the *Main* window:

16. Double-click on the base fingerprint type in the *Experiments* panel to open the *Fingerprint type* window.

Fingerp	orints						
Ð.							
No b	nds on highlighted curve (7 and selected	7962 pts)					
					-110	-120 -130 -130 -170	-190
	006-0003 [MLVAMP16						
	006-0003 [MLVAMP1V			L DA L		-	
	006-0003 [MLVAMP1N 006-0003 [MLVAMP1P					_	
_ ∎ N2	006-0003 [MLVAMP26	-FAM]					
[≫] ∎ N2	006-0003 [MLVAMP26 006-0003 [MLVAMP2V 006-0003 [MLVAMP2V	(IC]					
1112	006-0003 [MLVAMP2N 006-0003 [MLVAMP2P	coj		N h i			
112	000-0003 (in: v Aim 21			maller 1		I I	
N2	006-0004 [MLVAMP16	-FAM]			-		
N2	006-0004 [MLVAMP1V	/IC]					
	006-0004 [MLVAMP1N			l All A.	i I		
IN2	006-0004 [MLVAMP1P	E1]		LTL. MA			
				A A A A A A A A A A A A A A A A A A A	l l		
∎ N2	006-0004 [MLVAMP26	-FAM]	<			R	>
Channe	els			Overview			
8							
	File name	Channel	-				
6 🛛 🔇	Batch 1_LIZ	LIZ	^				
	Batch 1_6-FAM	6-FAM		- Weight - Constant -			
0	Batch 1_VIC	VIC					
	Batch 1_NED	NED		3			
	Batch 1_PET	PET					
1							

Figure 11: Normalized view - data dyes.

17. In the *Fingerprint type* window, call *Settings* > *Edit reference system*, or double-click in the *R01 panel* to call the *Fingerprint Reference system* window.

A calibration curve for the reference system **R01** of the base fingerprint type is displayed.

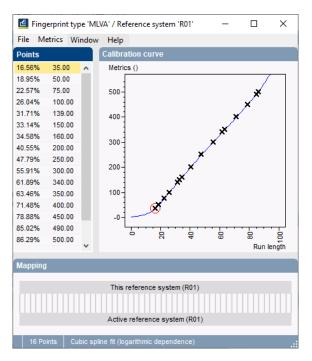


Figure 12: The *Fingerprint Reference system* window, with a calibration curve displayed.

18. Close the *Fingerprint Reference system* window and the *Fingerprint type* window.

5 Conclusion

In this tutorial you have seen how to import and process sequencer curve files in BIONUMERICS. You can now start applying comparison functions such as band matching, clustering, etc. on the data. More information about these functions can found in the analysis tutorials on our website.